1

Foundations of Matrix Analysis

In this chapter we recall the basic elements of linear algebra which will be employed in the remainder of the text. For most of the proofs as well as for the details, the reader is referred to [Bra75], [Nob69], [Hal58]. Further results on eigenvalues can be found in [Hou75] and [Wil65].

1.1 Vector Spaces

Definition 1.1 A *vector space* over the numeric field K ($K = \mathbb{R}$ or $K = \mathbb{C}$) is a nonempty set V, whose elements are called *vectors* and in which two operations are defined, called *addition* and *scalar multiplication*, that enjoy the following properties:

1. addition is commutative and associative;
2. there exists an element $\mathbf{0} \in V$ (the *zero vector* or *null vector*) such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for each $\mathbf{v} \in V$;
3. $0 \cdot \mathbf{v} = \mathbf{0}$, $1 \cdot \mathbf{v} = \mathbf{v}$, for each $\mathbf{v} \in V$, where 0 and 1 are respectively the zero and the unity of K;
4. for each element $\mathbf{v} \in V$ there exists its opposite, $-\mathbf{v}$, in V such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$;
5. the following distributive properties hold

$$\forall \alpha \in K,\ \forall \mathbf{v}, \mathbf{w} \in V,\ \alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w},$$

$$\forall \alpha, \beta \in K,\ \forall \mathbf{v} \in V,\ (\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v};$$

6. the following associative property holds

$$\forall \alpha, \beta \in K,\ \forall \mathbf{v} \in V,\ (\alpha\beta)\mathbf{v} = \alpha(\beta\mathbf{v}).$$

■

Example 1.1 Remarkable instances of vector spaces are:

- $V = \mathbb{R}^n$ (respectively $V = \mathbb{C}^n$): the set of the n-tuples of real (respectively complex) numbers, $n \geq 1$;
- $V = \mathbb{P}_n$: the set of polynomials $p_n(x) = \sum_{k=0}^{n} a_k x^k$ with real (or complex) coefficients a_k having degree less than or equal to n, $n \geq 0$;
- $V = C^p([a, b])$: the set of real (or complex)-valued functions which are continuous on $[a, b]$ up to their p-th derivative, $0 \leq p < \infty$. •

Definition 1.2 We say that a nonempty part W of V is a *vector subspace* of V iff W is a vector space over K. ■

Example 1.2 The vector space $\mathbb{P}_n$ is a vector subspace of $C^\infty(\mathbb{R})$, which is the space of infinite continuously differentiable functions on the real line. A trivial subspace of any vector space is the one containing only the zero vector. •

In particular, the set W of the linear combinations of a system of p vectors of V, $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, is a vector subspace of V, called the *generated subspace* or *span* of the vector system, and is denoted by

$$\begin{aligned} W &= \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} \\ &= \{\mathbf{v} = \alpha_1 \mathbf{v}_1 + \ldots + \alpha_p \mathbf{v}_p \quad \text{with } \alpha_i \in K,\ i = 1, \ldots, p\}. \end{aligned}$$

The system $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is called a system of *generators* for W.
If $W_1, \ldots, W_m$ are vector subspaces of V, then the set

$$S = \{\mathbf{w} : \ \mathbf{w} = \mathbf{v}_1 + \ldots + \mathbf{v}_m \text{ with } \mathbf{v}_i \in W_i,\ i = 1, \ldots, m\}$$

is also a vector subspace of V. We say that S is the *direct sum* of the subspaces W_i if any element $\mathbf{s} \in S$ admits a unique representation of the form $\mathbf{s} = \mathbf{v}_1 + \ldots + \mathbf{v}_m$ with $\mathbf{v}_i \in W_i$ and $i = 1, \ldots, m$. In such a case, we shall write $S = W_1 \oplus \ldots \oplus W_m$.

Definition 1.3 A system of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ of a vector space V is called *linearly independent* if the relation

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \ldots + \alpha_m \mathbf{v}_m = \mathbf{0}$$

with $\alpha_1, \alpha_2, \ldots, \alpha_m \in K$ implies that $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$. Otherwise, the system will be called *linearly dependent*. ■

We call a *basis* of V any system of linearly independent generators of V. If $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of V, the expression $\mathbf{v} = v_1 \mathbf{u}_1 + \ldots + v_n \mathbf{u}_n$ is called the *decomposition* of $\mathbf{v}$ with respect to the basis and the scalars $v_1, \ldots, v_n \in K$ are the *components* of $\mathbf{v}$ with respect to the given basis. Moreover, the following property holds.

Property 1.1 *Let V be a vector space which admits a basis of n vectors. Then every system of linearly independent vectors of V has at most n elements and any other basis of V has n elements. The number n is called the dimension of V and we write $dim(V) = n$.*
If, instead, for any n there always exist n linearly independent vectors of V, the vector space is called infinite dimensional.

Example 1.3 For any integer p the space $C^p([a, b])$ is infinite dimensional. The spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ have dimension equal to n. The usual basis for $\mathbb{R}^n$ is the set of *unit vectors* $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ where $(\mathbf{e}_i)_j = \delta_{ij}$ for $i, j = 1, \dots n$, where δ_{ij} denotes the *Kronecker symbol* equal to 0 if $i \neq j$ and 1 if $i = j$. This choice is of course not the only one that is possible (see Exercise 2). •

1.2 Matrices

Let m and n be two positive integers. We call a *matrix* having m rows and n columns, or a matrix $m \times n$, or a matrix (m, n), with elements in K, a set of mn scalars $a_{ij} \in K$, with $i = 1, \dots, m$ and $j = 1, \dots n$, represented in the following rectangular array

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}. \tag{1.1}$$

When $K = \mathbb{R}$ or $K = \mathbb{C}$ we shall respectively write $\mathrm{A} \in \mathbb{R}^{m \times n}$ or $\mathrm{A} \in \mathbb{C}^{m \times n}$, to explicitly outline the numerical fields which the elements of A belong to. Capital letters will be used to denote the matrices, while the lower case letters corresponding to those upper case letters will denote the matrix entries.
We shall abbreviate (1.1) as $\mathrm{A} = (a_{ij})$ with $i = 1, \dots, m$ and $j = 1, \dots n$. The index i is called row index, while j is the column index. The set $(a_{i1}, a_{i2}, \dots, a_{in})$ is called the *i-th row* of A; likewise, $(a_{1j}, a_{2j}, \dots, a_{mj})$ is the *j-th column* of A.

If $n = m$ the matrix is called *squared* or having order n and the set of the entries $(a_{11}, a_{22}, \dots, a_{nn})$ is called its *main diagonal*.

A matrix having one row or one column is called a *row vector* or *column vector* respectively. Unless otherwise specified, we shall always assume that a vector is a column vector. In the case $n = m = 1$, the matrix will simply denote a scalar of K.
Sometimes it turns out to be useful to distinguish within a matrix the set made up by specified rows and columns. This prompts us to introduce the following definition.

Definition 1.4 Let A be a matrix $m \times n$. Let $1 \leq i_1 < i_2 < \dots < i_k \leq m$ and $1 \leq j_1 < j_2 < \dots < j_l \leq n$ two sets of contiguous indexes. The matrix $\mathrm{S}(k \times l)$

of entries $s_{pq} = a_{i_p j_q}$ with $p = 1, \ldots, k$, $q = 1, \ldots, l$ is called a *submatrix* of A. If $k = l$ and $i_r = j_r$ for $r = 1, \ldots, k$, S is called a *principal submatrix* of A. ■

Definition 1.5 A matrix $\mathrm{A}(m \times n)$ is called *block partitioned* or said to be *partitioned into submatrices* if

$$\mathrm{A} = \begin{bmatrix} \mathrm{A}_{11} & \mathrm{A}_{12} & \ldots & \mathrm{A}_{1l} \\ \mathrm{A}_{21} & \mathrm{A}_{22} & \ldots & \mathrm{A}_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{A}_{k1} & \mathrm{A}_{k2} & \ldots & \mathrm{A}_{kl} \end{bmatrix},$$

where A_{ij} are submatrices of A. ■

Among the possible partitions of A, we recall in particular the partition by columns

$$\mathrm{A} = (\mathbf{a}_1,\ \mathbf{a}_2,\ \ldots, \mathbf{a}_n),$$

$\mathbf{a}_i$ being the i-th column vector of A. In a similar way the partition by rows of A can be defined. To fix the notations, if A is a matrix $m \times n$, we shall denote by

$$\mathrm{A}(i_1 : i_2, j_1 : j_2) = (a_{ij})\ i_1 \le i \le i_2,\ \ j_1 \le j \le j_2$$

the submatrix of A of size $(i_2 - i_1 + 1) \times (j_2 - j_1 + 1)$ that lies between the rows i_1 and i_2 and the columns j_1 and j_2. Likewise, if $\mathbf{v}$ is a vector of size n, we shall denote by $\mathbf{v}(i_1 : i_2)$ the vector of size $i_2 - i_1 + 1$ made up by the i_1-th to the i_2-th components of $\mathbf{v}$.

These notations are convenient in view of programming the algorithms that will be presented throughout the volume in the MATLAB language.

1.3 Operations with Matrices

Let $\mathrm{A} = (a_{ij})$ and $\mathrm{B} = (b_{ij})$ be two matrices $m \times n$ over K. We say that A is *equal* to B, if $a_{ij} = b_{ij}$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$. Moreover, we define the following operations:

- *matrix sum*: the matrix sum is the matrix $\mathrm{A} + \mathrm{B} = (a_{ij} + b_{ij})$. The neutral element in a matrix sum is the *null matrix*, still denoted by 0 and made up only by null entries;
- *matrix multiplication by a scalar*: the multiplication of A by $\lambda \in K$, is a matrix $\lambda\mathrm{A} = (\lambda a_{ij})$;
- *matrix product*: the product of two matrices A and B of sizes (m, p) and (p, n) respectively, is a matrix $\mathrm{C}(m, n)$ whose entries are $c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$, for $i = 1, \ldots, m$, $j = 1, \ldots, n$.

The matrix product is associative and distributive with respect to the matrix sum, but it is not in general commutative. The square matrices for which the property $AB = BA$ holds, will be called *commutative*.

In the case of square matrices, the neutral element in the matrix product is a square matrix of order n called the *unit matrix of order* n or, more frequently, the *identity matrix* given by $I_n = (\delta_{ij})$. The identity matrix is, by definition, the only matrix $n \times n$ such that $AI_n = I_n A = A$ for all square matrices A. In the following we shall omit the subscript n unless it is strictly necessary. The identity matrix is a special instance of a *diagonal matrix* of order n, that is, a square matrix of the type $D = (d_{ii}\delta_{ij})$. We will use in the following the notation $D = \mathrm{diag}(d_{11}, d_{22}, \ldots, d_{nn})$.
Finally, if A is a square matrix of order n and p is an integer, we define A^p as the product of A with itself iterated p times. We let $A^0 = I$.
Let us now address the so-called *elementary row operations* that can be performed on a matrix. They consist of:

- multiplying the i-th row of a matrix by a scalar α; this operation is equivalent to pre-multiplying A by the matrix $D = \mathrm{diag}(1, \ldots, 1, \alpha, 1, \ldots, 1)$, where α occupies the i-th position;
- exchanging the i-th and j-th rows of a matrix; this can be done by pre-multiplying A by the matrix $P^{(i,j)}$ of elements

$$p_{rs}^{(i,j)} = \begin{cases} 1 & \text{if } r = s = 1, \ldots, i-1, i+1, \ldots, j-1, j+1, \ldots n, \\ 1 & \text{if } r = j, s = i \text{ or } r = i, s = j, \\ 0 & \text{otherwise}. \end{cases} \tag{1.2}$$

  Matrices like (1.2) are called *elementary permutation matrices*. The product of elementary permutation matrices is called a *permutation matrix*, and it performs the row exchanges associated with each elementary permutation matrix. In practice, a permutation matrix is a reordering by rows of the identity matrix;
- adding α times the j-th row of a matrix to its i-th row. This operation can also be performed by pre-multiplying A by the matrix $I + N_\alpha^{(i,j)}$, where $N_\alpha^{(i,j)}$ is a matrix having null entries except the one in position i, j whose value is α.

1.3.1 Inverse of a Matrix

Definition 1.6 A square matrix A of order n is called *invertible* (or *regular* or *nonsingular*) if there exists a square matrix B of order n such that $A\,B = B\,A = I$. B is called the *inverse matrix* of A and is denoted by A^{-1}. A matrix which is not invertible is called *singular*. ■

If A is invertible its inverse is also invertible, with $(A^{-1})^{-1} = A$. Moreover, if A and B are two invertible matrices of order n, their product AB is also invertible, with $(A\,B)^{-1} = B^{-1}A^{-1}$. The following property holds.

Property 1.2 *A square matrix is invertible iff its column vectors are linearly independent.*

Definition 1.7 We call the *transpose* of a matrix $\mathrm{A} \in \mathbb{R}^{m\times n}$ the matrix $n\times m$, denoted by A^T, that is obtained by exchanging the rows of A with the columns of A. ■

Clearly, $(\mathrm{A}^T)^T = \mathrm{A}$, $(\mathrm{A}+\mathrm{B})^T = \mathrm{A}^T+\mathrm{B}^T$, $(\mathrm{AB})^T = \mathrm{B}^T\mathrm{A}^T$ and $(\alpha\mathrm{A})^T = \alpha\mathrm{A}^T$ $\forall\alpha \in \mathbb{R}$. If A is invertible, then also $(\mathrm{A}^T)^{-1} = (\mathrm{A}^{-1})^T = \mathrm{A}^{-T}$.

Definition 1.8 Let $\mathrm{A} \in \mathbb{C}^{m\times n}$; the matrix $\mathrm{B} = \mathrm{A}^H \in \mathbb{C}^{n\times m}$ is called the *conjugate transpose* (or *adjoint*) of A if $b_{ij} = \bar{a}_{ji}$, where $\bar{a}_{ji}$ is the complex conjugate of a_{ji}. ■

In analogy with the case of the real matrices, it turns out that $(\mathrm{A}+\mathrm{B})^H = \mathrm{A}^H + \mathrm{B}^H$, $(\mathrm{AB})^H = \mathrm{B}^H\mathrm{A}^H$ and $(\alpha\mathrm{A})^H = \bar{\alpha}\mathrm{A}^H$ $\forall\alpha \in \mathbb{C}$.

Definition 1.9 A matrix $\mathrm{A} \in \mathbb{R}^{n\times n}$ is called *symmetric* if $\mathrm{A} = \mathrm{A}^T$, while it is *antisymmetric* if $\mathrm{A} = -\mathrm{A}^T$. Finally, it is called *orthogonal* if $\mathrm{A}^T\mathrm{A} = \mathrm{A}\mathrm{A}^T = \mathrm{I}$, that is $\mathrm{A}^{-1} = \mathrm{A}^T$. ■

Permutation matrices are orthogonal and the same is true for their products.

Definition 1.10 A matrix $\mathrm{A} \in \mathbb{C}^{n\times n}$ is called *hermitian* or *self-adjoint* if $\mathrm{A}^T = \bar{\mathrm{A}}$, that is, if $\mathrm{A}^H = \mathrm{A}$, while it is called *unitary* if $\mathrm{A}^H\mathrm{A} = \mathrm{A}\mathrm{A}^H = \mathrm{I}$. Finally, if $\mathrm{A}\mathrm{A}^H = \mathrm{A}^H\mathrm{A}$, A is called *normal*. ■

As a consequence, a unitary matrix is one such that $\mathrm{A}^{-1} = \mathrm{A}^H$.
Of course, a unitary matrix is also normal, but it is not in general hermitian. For instance, the matrix of the Example 1.4 is unitary, although not symmetric (if $s \neq 0$). We finally notice that the diagonal entries of an hermitian matrix must necessarily be real (see also Exercise 5).

1.3.2 Matrices and Linear Mappings

Definition 1.11 A *linear map* from $\mathbb{C}^n$ into $\mathbb{C}^m$ is a function $f : \mathbb{C}^n \longrightarrow \mathbb{C}^m$ such that $f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$, $\forall\alpha, \beta \in K$ and $\forall\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. ■

The following result links matrices and linear maps.

Property 1.3 *Let $f : \mathbb{C}^n \longrightarrow \mathbb{C}^m$ be a linear map. Then, there exists a unique matrix $\mathrm{A}_f \in \mathbb{C}^{m\times n}$ such that*

$$f(\mathbf{x}) = \mathrm{A}_f\mathbf{x} \qquad \forall\mathbf{x} \in \mathbb{C}^n. \tag{1.3}$$

Conversely, if $\mathrm{A}_f \in \mathbb{C}^{m\times n}$ then the function defined in (1.3) *is a linear map from $\mathbb{C}^n$ into $\mathbb{C}^m$.*

Example 1.4 An important example of a linear map is the counterclockwise *rotation* by an angle ϑ in the plane (x_1, x_2). The matrix associated with such a map is given by

$$\mathrm{G}(\vartheta) = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}, \quad c = \cos(\vartheta), \;\; s = \sin(\vartheta)$$

and it is called a *rotation matrix*. •

1.3.3 Operations with Block-Partitioned Matrices

All the operations that have been previously introduced can be extended to the case of a block-partitioned matrix A, provided that the size of each single block is such that any single matrix operation is well-defined.

Indeed, the following result can be shown (see, e.g., [Ste73]).

Property 1.4 *Let* A *and* B *be the block matrices*

$$\mathrm{A} = \begin{bmatrix} \mathrm{A}_{11} & \dots & \mathrm{A}_{1l} \\ \vdots & \ddots & \vdots \\ \mathrm{A}_{k1} & \dots & \mathrm{A}_{kl} \end{bmatrix}, \mathrm{B} = \begin{bmatrix} \mathrm{B}_{11} & \dots & \mathrm{B}_{1n} \\ \vdots & \ddots & \vdots \\ \mathrm{B}_{m1} & \dots & \mathrm{B}_{mn} \end{bmatrix},$$

where A_{ij} *and* B_{ij} *are matrices* $(k_i \times l_j)$ *and* $(m_i \times n_j)$*. Then we have*

1.

$$\lambda\mathrm{A} = \begin{bmatrix} \lambda\mathrm{A}_{11} & \dots & \lambda\mathrm{A}_{1l} \\ \vdots & \ddots & \vdots \\ \lambda\mathrm{A}_{k1} & \dots & \lambda\mathrm{A}_{kl} \end{bmatrix}, \quad \lambda \in \mathbb{C};\ \mathrm{A}^T = \begin{bmatrix} \mathrm{A}_{11}^T & \dots & \mathrm{A}_{k1}^T \\ \vdots & \ddots & \vdots \\ \mathrm{A}_{1l}^T & \dots & \mathrm{A}_{kl}^T \end{bmatrix};$$

2. *if* $k = m$, $l = n$, $m_i = k_i$ *and* $n_j = l_j$, *then*

$$\mathrm{A} + \mathrm{B} = \begin{bmatrix} \mathrm{A}_{11} + \mathrm{B}_{11} & \dots & \mathrm{A}_{1l} + \mathrm{B}_{1l} \\ \vdots & \ddots & \vdots \\ \mathrm{A}_{k1} + \mathrm{B}_{k1} & \dots & \mathrm{A}_{kl} + \mathrm{B}_{kl} \end{bmatrix};$$

3. *if* $l = m$, $l_i = m_i$ *and* $k_i = n_i$, *then, letting* $C_{ij} = \sum_{s=1}^{m} \mathrm{A}_{is}\mathrm{B}_{sj}$,

$$\mathrm{AB} = \begin{bmatrix} \mathrm{C}_{11} & \dots & \mathrm{C}_{1l} \\ \vdots & \ddots & \vdots \\ \mathrm{C}_{k1} & \dots & \mathrm{C}_{kl} \end{bmatrix}.$$

1.4 Trace and Determinant of a Matrix

Let us consider a square matrix A of order n. The *trace* of a matrix is the sum of the diagonal entries of A, that is $\mathrm{tr}(\mathrm{A}) = \sum_{i=1}^{n} a_{ii}$.
We call the *determinant* of A the scalar defined through the following formula

$$\det(\mathrm{A}) = \sum_{\boldsymbol{\pi} \in P} \mathrm{sign}(\boldsymbol{\pi}) a_{1\pi_1} a_{2\pi_2} \dots a_{n\pi_n},$$

where $P = \{\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)^T\}$ is the set of the $n!$ vectors that are obtained by permuting the index vector $\mathbf{i} = (1, \dots, n)^T$ and $\mathrm{sign}(\boldsymbol{\pi})$ equal to 1 (respectively, -1) if an even (respectively, odd) number of exchanges is needed to obtain $\boldsymbol{\pi}$ from $\mathbf{i}$.
The following properties hold

$$\det(\mathrm{A}) = \det(\mathrm{A}^T),\ \det(\mathrm{AB}) = \det(\mathrm{A})\det(\mathrm{B}),\ \det(\mathrm{A}^{-1}) = 1/\det(\mathrm{A}),$$

$$\det(\mathrm{A}^H) = \overline{\det(\mathrm{A})},\ \det(\alpha\mathrm{A}) = \alpha^n \det(\mathrm{A}),\ \forall \alpha \in K.$$

Moreover, if two rows or columns of a matrix coincide, the determinant vanishes, while exchanging two rows (or two columns) produces a change of sign in the determinant. Of course, the determinant of a diagonal matrix is the product of the diagonal entries.

Denoting by A_{ij} the matrix of order $n-1$ obtained from A by eliminating the i-th row and the j-th column, we call the *complementary minor* associated with the entry a_{ij} the determinant of the matrix A_{ij}. We call the *k-th principal (dominating) minor* of A, d_k, the determinant of the principal submatrix of order k, $\mathrm{A}_k = \mathrm{A}(1:k, 1:k)$. If we denote by $\Delta_{ij} = (-1)^{i+j}\det(\mathrm{A}_{ij})$ the *cofactor* of the entry a_{ij}, the actual computation of the determinant of A can be performed using the following recursive relation

$$\det(\mathrm{A}) = \begin{cases} a_{11} & \text{if } n = 1, \\ \\ \displaystyle\sum_{j=1}^{n} \Delta_{ij} a_{ij}, & \text{for } n > 1, \end{cases} \tag{1.4}$$

which is known as the *Laplace rule*. If A is a square invertible matrix of order n, then

$$\mathrm{A}^{-1} = \frac{1}{\det(\mathrm{A})} \mathrm{C},$$

where C is the matrix having entries Δ_{ji}, $i, j = 1, \dots, n$.

As a consequence, a square matrix is invertible iff its determinant is nonvanishing. In the case of nonsingular diagonal matrices the inverse is still a diagonal matrix having entries given by the reciprocals of the diagonal entries of the matrix.

Every *orthogonal matrix* is invertible, its inverse is given by A^T, moreover $\det(\mathrm{A}) = \pm 1$.

1.5 Rank and Kernel of a Matrix

Let A be a rectangular matrix $m \times n$. We call the *determinant of order q (with $q \geq 1$) extracted from matrix* A, the determinant of any square matrix of order q obtained from A by eliminating $m - q$ rows and $n - q$ columns.

Definition 1.12 The *rank* of A (denoted by rank(A)) is the maximum order of the nonvanishing determinants extracted from A. A matrix has *complete or full rank* if $\mathrm{rank}(\mathrm{A}) = \min(m,n)$. ■

Notice that the rank of A represents the maximum number of linearly independent column vectors of A that is, the dimension of the *range* of A, defined as

$$\mathrm{range}(\mathrm{A}) = \{\mathbf{y} \in \mathbb{R}^m : \ \mathbf{y} = \mathrm{A}\mathbf{x} \text{ for } \mathbf{x} \in \mathbb{R}^n\}\,. \tag{1.5}$$

Rigorously speaking, one should distinguish between the column rank of A and the row rank of A, the latter being the maximum number of linearly independent row vectors of A. Nevertheless, it can be shown that the row rank and column rank do actually coincide.

The *kernel* of A is defined as the subspace

$$\ker(\mathrm{A}) = \{\mathbf{x} \in \mathbb{R}^n : \ \mathrm{A}\mathbf{x} = \mathbf{0}\}\,.$$

The following relations hold:

1. $\mathrm{rank}(\mathrm{A}) = \mathrm{rank}(\mathrm{A}^T) \quad (\text{if } \mathrm{A} \in \mathbb{C}^{m\times n},\ \mathrm{rank}(\mathrm{A}) = \mathrm{rank}(\mathrm{A}^H))$;
2. $\mathrm{rank}(\mathrm{A}) + \dim(\ker(\mathrm{A})) = n$.

In general, $\dim(\ker(\mathrm{A})) \neq \dim(\ker(\mathrm{A}^T))$. If A is a nonsingular square matrix, then $\mathrm{rank}(\mathrm{A}) = n$ and $\dim(\ker(\mathrm{A})) = 0$.

Example 1.5 Let

$$\mathrm{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}.$$

Then, $\mathrm{rank}(\mathrm{A}) = 2$, $\dim(\ker(\mathrm{A})) = 1$ and $\dim(\ker(\mathrm{A}^T)) = 0$. •

We finally notice that for a matrix $A \in \mathbb{C}^{n\times n}$ the following properties are equivalent:

1. A is nonsingular;
2. $\det(A) \neq 0$;
3. $\ker(A) = \{\mathbf{0}\}$;
4. $\text{rank}(A) = n$;
5. A has linearly independent rows and columns.

1.6 Special Matrices

1.6.1 Block Diagonal Matrices

These are matrices of the form $D = \text{diag}(D_1, \ldots, D_n)$, where D_i are square matrices with $i = 1, \ldots, n$. Clearly, each single diagonal block can be of different size. We shall say that a block diagonal matrix has size n if n is the number of its diagonal blocks. The determinant of a block diagonal matrix is given by the product of the determinants of the single diagonal blocks.

1.6.2 Trapezoidal and Triangular Matrices

A matrix $A(m \times n)$ is called *upper trapezoidal* if $a_{ij} = 0$ for $i > j$, while it is *lower trapezoidal* if $a_{ij} = 0$ for $i < j$. The name is due to the fact that, in the case of upper trapezoidal matrices, with $m < n$, the nonzero entries of the matrix form a trapezoid.

A *triangular matrix* is a square trapezoidal matrix of order n of the form

$$L = \begin{bmatrix} l_{11} & 0 & \ldots & 0 \\ l_{21} & l_{22} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ l_{n1} & l_{n2} & \ldots & l_{nn} \end{bmatrix} \text{ or } U = \begin{bmatrix} u_{11} & u_{12} & \ldots & u_{1n} \\ 0 & u_{22} & \ldots & u_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & u_{nn} \end{bmatrix}.$$

The matrix L is called *lower triangular* while U is *upper triangular*.
Let us recall some algebraic properties of triangular matrices that are easy to check.

- The determinant of a triangular matrix is the product of the diagonal entries;
- the inverse of a lower (respectively, upper) triangular matrix is still lower (respectively, upper) triangular;
- the product of two lower triangular (respectively, upper trapezoidal) matrices is still lower triangular (respectively, upper trapezoidal);
- if we call *unit triangular matrix* a triangular matrix that has diagonal entries equal to 1, then, the product of lower (respectively, upper) unit triangular matrices is still lower (respectively, upper) unit triangular.

1.6.3 Banded Matrices

The matrices introduced in the previous section are a special instance of banded matrices. Indeed, we say that a matrix $\mathrm{A} \in \mathbb{R}^{m\times n}$ (or in $\mathbb{C}^{m\times n}$) has *lower band* p if $a_{ij} = 0$ when $i > j + p$ and *upper band* q if $a_{ij} = 0$ when $j > i + q$. Diagonal matrices are banded matrices for which $p = q = 0$, while trapezoidal matrices have $p = m - 1$, $q = 0$ (lower trapezoidal), $p = 0$, $q = n - 1$ (upper trapezoidal).

Other banded matrices of relevant interest are the *tridiagonal matrices* for which $p = q = 1$ and the *upper bidiagonal* ($p = 0$, $q = 1$) or *lower bidiagonal* ($p = 1$, $q = 0$). In the following, $\mathrm{tridiag}_n(\mathbf{b}, \mathbf{d}, \mathbf{c})$ will denote the triadiagonal matrix of size n having respectively on the lower and upper principal diagonals the vectors $\mathbf{b} = (b_1, \ldots, b_{n-1})^T$ and $\mathbf{c} = (c_1, \ldots, c_{n-1})^T$, and on the principal diagonal the vector $\mathbf{d} = (d_1, \ldots, d_n)^T$. If $b_i = \beta$, $d_i = \delta$ and $c_i = \gamma$, β, δ and γ being given constants, the matrix will be denoted by $\mathrm{tridiag}_n(\beta, \delta, \gamma)$.

We also mention the so-called *lower Hessenberg matrices* ($p = m - 1$, $q = 1$) and *upper Hessenberg matrices* ($p = 1$, $q = n - 1$) that have the following structure

$$\mathrm{H} = \begin{bmatrix} h_{11} & h_{12} & & \text{\Large 0} \\ h_{21} & h_{22} & \ddots & \\ \vdots & & \ddots & h_{m-1n} \\ h_{m1} & \ldots & \ldots & h_{mn} \end{bmatrix} \text{ or } \mathrm{H} = \begin{bmatrix} h_{11} & h_{12} & \ldots & h_{1n} \\ h_{21} & h_{22} & & h_{2n} \\ & \ddots & \ddots & \vdots \\ \text{\Large 0} & & h_{mn-1} & h_{mn} \end{bmatrix}.$$

Matrices of similar shape can obviously be set up in the block-like format.

1.7 Eigenvalues and Eigenvectors

Let A be a square matrix of order n with real or complex entries; the number $\lambda \in \mathbb{C}$ is called an *eigenvalue* of A if there exists a nonnull vector $\mathbf{x} \in \mathbb{C}^n$ such that $\mathrm{A}\mathbf{x} = \lambda\mathbf{x}$. The vector $\mathbf{x}$ is the *eigenvector* associated with the eigenvalue λ and the set of the eigenvalues of A is called the *spectrum* of A, denoted by $\sigma(\mathrm{A})$. We say that $\mathbf{x}$ and $\mathbf{y}$ are respectively a *right eigenvector* and a *left eigenvector* of A, associated with the eigenvalue λ, if

$$\mathrm{A}\mathbf{x} = \lambda\mathbf{x},\ \mathbf{y}^H\mathrm{A} = \lambda\mathbf{y}^H.$$

The eigenvalue λ corresponding to the eigenvector $\mathbf{x}$ can be determined by computing the *Rayleigh quotient* $\lambda = \mathbf{x}^H\mathrm{A}\mathbf{x}/(\mathbf{x}^H\mathbf{x})$. The number λ is the solution of the *characteristic equation*

$$p_{\mathrm{A}}(\lambda) = \det(\mathrm{A} - \lambda\mathrm{I}) = 0,$$

where $p_{\mathrm{A}}(\lambda)$ is the *characteristic polynomial*. Since this latter is a polynomial of degree n with respect to λ, there certainly exist n eigenvalues of A not necessarily distinct. The following properties can be proved